Optimal convergence rates for the invariant density estimation of jump-diffusion processes

Chiara Amorino*and Eulalia Nualart†

January 14, 2021

Abstract

We aim at estimating the invariant density associated to a stochastic differential equation with jumps in low dimension, which is for $d = 1$ and $d = 2$. We consider a class of jump diffusion processes whose invariant density belongs to some Hölder space. Firstly, in dimension one, we show that the kernel density estimator achieves the convergence rate $\frac{1}{T}$, which is the optimal rate in the absence of jumps. This improves the convergence rate obtained in [2], which depends on the Blumenthal-Getoor index for $d = 1$ and is equal to $\frac{\log T}{T}$ for $d = 2$. Secondly, we show that is not possible to find an estimator with faster rates of estimation. Indeed, we get some lower bounds with the same rates $\{\frac{1}{T}, \frac{\log T}{T}\}$ in the mono and bi-dimensional cases, respectively. Finally, we obtain the asymptotic normality of the estimator in the one-dimensional case.

Keywords: Minimax risk, convergence rate, non-parametric statistics, ergodic diffusion with jumps, Lévy driven SDE, invariant density estimation

1 Introduction

Solutions to Lévy-driven stochastic differential equations have recently attracted a lot of attention in the literature due to its many applications in various areas such as finance, physics, and neuroscience. Indeed, it includes some important examples from finance such as the well-known Kou model in [29], the Barndorff-Nielsen-Shephard model ([8]), and the Merton model ([32]) to name just a few. An important example of application of jump-processes in neuroscience is the stochastic Morris-Lecar neuron model presented in [25]. As a consequence, statistical inference for jump processes has recently become an active domain of research.

*Université du Luxembourg, L-4364 Esch-Sur-Alzette, Luxembourg. CA gratefully acknowledges financial support of ERC Consolidator Grant 815703 STAMFORD: Statistical Methods for High Dimensional Diffusions.

†Universitat Pompeu Fabra and Barcelona Graduate School of Economics, Department of Economics and Business, Ramón Trias Fargas 25-27, 08005 Barcelona, Spain. EN acknowledges support from the Spanish MINECO grant PGC2018-101643-B-I00 and Ayudas Fundacion BBVA a Equipos de Investigación Científica 2017.

We consider the process $(X_t)_{t\geq 0}$ solution to the following stochastic differential equation with jumps:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dB_s + \int_0^t \int_{\mathbb{R}_0^d} \gamma(X_{s^-})z(\nu(ds,dz) - F(z)dzds), \quad (1)$$

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion and ν is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ associated to a Lévy process $(L_t)_{t\geq 0}$ with Lévy density function F. We focus on the estimation of the invariant density μ associated to the jump-process solution to (1) in low dimension, which is for $d = 1$ and $d = 2$. In particular, assuming that a continuous record of $(X_t)_{t\in[0,T]}$ is available, our goal is to propose a non-parametric kernel estimator for the estimation of the stationary measure and to discuss its convergence rate for large T.

The same framework has been considered in some recent papers such as [2], [23] (Section 5.2), and [3]. In the first paper, it is shown that the kernel estimator achieves the following convergence rates for the pointwise estimation of the invariant density: $\frac{\log T}{T}$ for $d = 2$ and $\frac{(\log T)^{(2-\frac{(1+\alpha)}{2})\vee 1}}{T}$ for $d = 1$ (where α is the Blumenthal-Getoor index). We recall that, in the absence of jumps, the optimal convergence rate in the one-dimensional case is $\frac{1}{T}$, while the one found in [2] depends on the jumps and belongs to the interval $(\frac{\log T}{T}, \frac{(\log T)^{\frac{3}{2}}}{T})$.

In this paper, we wonder if such a deterioration on the rate is because of the presence of jumps or the used approach. Indeed, our purpose is to look for a new approach to recover a better convergence rate in the one-dimensional case (hopefully the same as in the continuous case) and to discuss the optimality of such a rate. This new approach will also lead to the obtaining of the asymptotic normality of the proposed estimator. After that, we will discuss the optimality of the convergence rate in the bi-dimensional case. This will close the circle of the analysis of the convergence rates for the estimation of the invariant density of jump-diffusions, as the convergence rates and their optimality in the case $d \geq 3$ have already been treated in detail in [3].

Beyond these works, to our best knowledge, the literature concerning non-parametric estimation of diffusion processes with jumps is not wide. One of the few examples is given by Funke and Schmisser: in [27] they investigate the non parametric adaptive estimation of the drift of an integrated jump diffusion process, while in [35], Schmisser deals with the non-parametric adaptive estimation of the coefficients of a jumps diffusion process. To name other examples, in [24] the authors estimate in a non-parametric way the drift of a diffusion with jumps driven by a Hawkes process, while in [4] the volatility and the jump coefficients are considered.

On the other hand, the problem of invariant density estimation has been considered by many authors (see e.g. [33], [20], [10], [39], and [5]) in several different frameworks: it is at the same time a long-standing problem and a highly active current topic of research. One of the reasons why the estimation of the invariant density has attracted the attention of many statisticians is the huge amount of numerical methods to which it is connected, the MCMC method above all. An approximation algorithm for the computation of the invariant density can be found for example in [30] and [34]. Moreover, invariant distributions are essential for the analysis of the stability of stochastic differential systems (see e.g. [28] and [5]).

In [5], [6], and [11] some kernel estimators are used to estimate the marginal density of a continuous time process. When μ belongs to some Hölder class whose smoothness is β, they prove under some mixing conditions that their pointwise L^2 risk achieves the standard rate of convergence $T^{\frac{2\beta}{2\beta+1}}$ and the rates are minimax in their framework. Castellana and Leadbetter proved in [15] that, under the following condition **CL**, the density can be estimated with the parametric rate $\frac{1}{T}$ by some non-parametric estimators (the kernel ones among them).

CL: $u \mapsto \|g_u\|_\infty$ is integrable on $(0, \infty)$ and $g_u(\cdot, \cdot)$ is continuous for each $u > 0$.

In our context, $g_u(x, y) = \mu(x)p_u(x, y) - \mu(x)\mu(y)$, where $p_u(x, y)$ is the transition density. More precisely, they shed light to the fact that local irregularities of the sample paths provide some additional information. Indeed, if the joint distribution of (X_0, X_t) is not too close to a singular distribution for $|t|$ small, then it is possible to achieve the superoptimal rate $\frac{1}{T}$ for the pointwise quadratic risk of the kernel estimator. Condition **CL** can be verified for ergodic continuous diffusion processes (see [38] for sufficient conditions). The paper of Castellana and Leadbetter led to a lot of works regarding the estimation of the common marginal distribution of a continuous time process. In [9], [10], [14], [21], and [7] several related results and examples can be found.

An alternative to the kernel density estimator is given by the local time density estimator, which was proposed by Kutoyants in [22] in the case of diffusion processes and was extended by Bosq and Davydov in [12] to a more general context. The latest have proved that, under a condition which is mildly weaker than **CL**, the mean squared error of the local time estimator reaches the full rate $\frac{1}{T}$. Leblanc built in [31] a wavelet estimator of a density belonging to some general Besov space and proved that, if the process is geometrically strong mixing and a condition like **CL** is satisfied, then its L^p-integrated risk converges at rate $\frac{1}{T}$ as well. In [18] the authors built a projection estimator and showed that its L^2-integrated risk achieves the parametric rate $\frac{1}{T}$ under a condition named WCL, which is blandly different compared to **CL**.

WCL: There exists a positive integrable function k (defined on $\mathbb{R}$) such that

$$\sup_{y\in\mathbb{R}} \int_0^\infty g_u(x, y)du \le k(x), \qquad \text{for all } x \in \mathbb{R}.$$

In this paper, we will show that our mono-dimensional jump-process satisfies a local irregularity condition **WCL1** and an asymptotic independence condition **WCL2** (see Proposition 1), two conditions in which the original condition **WCL** can be decomposed. In this way, it will be possible to show that the L^2 risk for the pointwise estimation of the invariant measure achieves the superoptimal rate $\frac{1}{T}$, using our kernel density estimator. Moreover, the same conditions will result in the asymptotic normality of the proposed estimator. Indeed, as we will see in the proof of Theorem 2, the main challenge in this part is to justify the use of dominated convergence theorem, which will ensured by conditions **WCL1** and **WCL2**. We will find in particular that, for any collection $(x_i)_{1\le i\le m}$ of real numbers, we have

$$\sqrt{T}(\hat{\mu}_{h,T}(x_i) - \mu(x_i),\, 1 \le i \le m) \xrightarrow{\mathcal{D}} N^{(m)}(0, \Sigma^{(m)}) \text{ as } T \to \infty,$$

where $\hat{\mu}_{h,T}$ is the kernel density estimator and

$$\Sigma^{(m)} := (\sigma(x_i, x_j))_{1\le i,j\le m}, \qquad \sigma(x_i, x_j) := 2\int_0^\infty g_u(x_i, x_j)du.$$

We remark that the precise form of the equation above allows us to construct tests and confidence sets for the density.

We have found the convergence rate $\left\{\frac{1}{T}, \frac{\log T}{T}\right\}$ for the risk associated to our kernel density estimator for the estimation of the invariant density for $d = 1$ and $d = 2$. Then, some questions naturally arise: are the convergence rates the best possible or is it possible to improve them by using other estimators? In order to answer, we consider a simpler model where both the volatility and the jump coefficient are constant and the intensity of the jumps is finite. Then, we look for a lower bound for the risk at a point $x \in \mathbb{R}^d$ defined as in equation (9) below. The first idea is to use the two hypothesis method (see Section 2.3 in [37]). To do that, the knowledge of the link between the drift b and the invariant density μ_b is essential. If in absence of jumps such link is explicit, in our context it is more challenging. As shown in [19] and [3], it is possible to find the link knowing that the invariant measure has to satisfy $A^*\mu_b = 0$, where A^* is the adjoint of the generator of the considered diffusion. This method allows us to show that the superoptimal rate $\frac{1}{T}$ is the best possible for the estimation of the invariant density in $d = 1$, but it fails in the bi-dimensional case (see Remark 1 below for details). Finally, we use a finite number of hypotheses to prove a lower bound in the bi-dimensional case. This requires a detailed analysis of the Kullback divergence between the probability laws associated to the different hypotheses. Thanks to that, it is possible to recover the optimal rate $\frac{\log T}{T}$ in the two-dimensional case.

The paper is organised as follows. In Section 2 we give the assumptions on our model and we provide our main results. Section 3 is devoted to state and prove some preliminary results needed for the proofs of the main results. To conclude, in Section 4 we give the proof of Theorems 1, 2, 3, and 4, where our main results are gathered.

2 Model assumption and main results

We consider the following stochastic differential equation with jumps

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dB_s + \int_0^t \int_{\mathbb{R}_0^d} \gamma(X_{s^-})z(\nu(ds, dz) - F(z)dzds), \quad (2)$$

where $t \ge 0$, $d \in \{1, 2\}$, $\mathbb{R}_0^d = \mathbb{R}^d \backslash \{0\}$, the initial condition X_0 is a $\mathbb{R}^d$-valued random variable, the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $\gamma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are measurable functions, $(B_t)_{t\ge 0}$ is a d-dimensional Brownian motion, and ν is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ associated to a Lévy process $(L_t)_{t\ge 0}$ with Lévy density function F. All sources of randomness are mutually independent.

We consider the following assumptions on the coefficients and on the Lévy density F:

A1 The functions $b(x)$, $\gamma(x)$ and $aa^T(x)$ are globally Lipschitz and bounded. Moreover, $\inf_{x\in\mathbb{R}} aa^T(x) \ge c\mathrm{Id}$, for some constant $c > 0$, where Id denotes the $d \times d$ identity matrix and $\inf_{x\in\mathbb{R}} \det(\gamma(x)) > 0$.

A2 $\langle x, b(x)\rangle \leq -c_1|x| + c_2$, for all $|x| \geq \rho$, for some $\rho, c_1, c_2 > 0$.

A3 $\text{Supp}(F) = \mathbb{R}_0^d$ and for all $z \in \mathbb{R}_0^d$, $F(z) \leq \frac{c_3}{|z|^{d+\alpha}}$, for some $\alpha \in (0,2), c_3 > 0$.

A4 There exist $\epsilon_0 > 0$ and $c_4 > 0$ such that $\int_{\mathbb{R}_0^d} |z|^2 e^{\epsilon_0|z|}F(z)dz \leq c_4$.

A5 If $\alpha = 1$, $\int_{r<|z|<R} zF(z)dz = 0$, for any $0 < r < R < \infty$.

Assumption **A1** ensures that equation (2) admits a unique càdlàg adapted solution $X = (X_t)_{t\geq 0}$ satisfying the strong Markov property, see e.g. [1]. Moreover, it is shown in [2, Lemma 2] that if we further assume Assumptions **A2-A4**, then the process X is exponentially ergodic and exponentially β-mixing. Therefore, it has a unique invariant distribution π, which we assume it has a density μ with respect to the Lebesgue measure. Finally, Assumption **A5** ensures the existence of the transition density of X denoted by $p_t(x,y)$ which satisfies the following upper bound (see [2, Lemma 1]): for all $T \geq 0$, there exists $c_0 > 0$ and $\lambda_0 > 0$ such that for any $t \in [0,T]$ and $x, y \in \mathbb{R}^d$,

$$p_t(x,y) \leq c_0 \left(t^{-d/2} e^{-\lambda_0 \frac{|y-x|^2}{t}} + \frac{t}{(t^{1/2} + |y-x|)^{d+\alpha}} \right). \tag{3}$$

We assume that the process is observed continuously $X = (X_t)_{t\in[0,T]}$ in a time interval $[0,T]$ such that T tends to ∞. In the paper [2] cited above, the nonparametric estimation of μ is studied via the kernel estimator which is defined as follows. We assume that μ belongs to the Hölder space $\mathcal{H}_d(\beta, \mathcal{L})$ where $\beta = (\beta_1, \ldots, \beta_d)$, $\beta_i > 1$ and $\mathcal{L} = (\mathcal{L}_1, \ldots, \mathcal{L}_d)$, $\mathcal{L}_i > 0$, which means that for all $i \in \{1, \ldots, d\}$, $k = 0, 1, \ldots, \lfloor \beta_i \rfloor$ and $t \in \mathbb{R}$,

$$\left\| D_i^{(k)} \mu \right\|_\infty \leq \mathcal{L} \quad \text{and} \quad \left\| D_i^{(\lfloor \beta_i \rfloor)} \mu(.+te_i) - D_i^{(\lfloor \beta_i \rfloor)} \mu(.) \right\|_\infty \leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor},$$

where $D_i^{(k)}$ denotes the kth order partial derivative of μ w.r.t the ith component, $\lfloor \beta_i \rfloor$ is the integer part of β_i, and $e_1, \ldots, e_d$ is the canonical basis of $\mathbb{R}^d$. We set

$$\hat{\mu}_{h,T}(x) = \frac{1}{T\prod_{i=1}^d h_i} \int_0^T \prod_{i=1}^d K\left(\frac{x_i - X_t^i}{h_i}\right) dt =: \frac{1}{T} \int_0^T \mathbb{K}_h(x - X_t)dt,$$

where $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $h = (h_1, \ldots, h_d)$ is a bandwidth and $K : \mathbb{R} \to \mathbb{R}$ is a kernel function satisfying

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \|K\|_\infty < \infty, \quad \text{supp}(K) \subset [-1,1], \quad \int_{\mathbb{R}} K(x)x^i dx = 0,$$

for all $i \in \{0, \ldots, M\}$ with $M \geq \max_i \beta_i$.

We first consider equation (2) with $d = 1$ and show that the kernel estimator reaches the optimal rate T^{-1}, as it is for the stochastic differential equation (2) without jumps. For this, we need the following additional assumption on F.

A6 F belongs to $C^1(\mathbb{R})$ and for all $z \in \mathbb{R}$, $|F'(z)| \leq \frac{c_5}{|z|^{2+\alpha}}$, for some $c_5 > 0$.

Theorem 1. *Let X be the solution to* (2) *on* $[0, T]$ *with* $d = 1$. *Suppose that Assumptions* **A1**-**A6** *hold and* $\mu \in \mathcal{H}_1(\beta, \mathcal{L})$. *Then there exists a constant* $c > 0$ *independent of* T *and* h *such that for all* $x \in \mathbb{R}$,

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \leq c(h^{2\beta} + \frac{1}{T}). \tag{4}$$

In particular, choosing $h(T) = \frac{1}{T^a}$ *with* $a > \frac{1}{2\beta}$, *we conclude that*

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \leq \frac{c}{T}.$$

Theorem 1 improves the upper bound obtained in [2] which was of the form $\frac{(\log T)^{(2-\frac{1+\alpha}{2})\vee 1}}{T}$. As in that paper, we will use the bias-variance decomposition (see [17, Proposition 1])

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \leq c\left(h^{2\beta} + T^{-2}\mathrm{Var}\left(\int_0^T \mathbb{K}(x - X_t)dt\right)\right). \tag{5}$$

Then in [2] bounds on the transition semigroup and on the transition density (see (3) above) give an upper bound for the variance depending on the bandwidth. Here, we use the same approach as in [15] and [18] to obtain a bandwidth-free rate for the variance of smoothing density estimators (which include the kernel estimator). For Markov diffusions, the sufficient conditions can be decomposed into a local irregularity condition **WCL1** plus an asymptotic independence condition **WCL2**:

$$\textbf{WCL1:} \quad \int_{\mathbb{R}} \int_0^1 \sup_{y\in\mathbb{R}} |g_u(x, y)|\, du\, dx < \infty,$$
$$\textbf{WCL2:} \quad \int_{\mathbb{R}} \int_1^\infty \sup_{y\in\mathbb{R}} |g_u(x, y)|\, du\, dx < \infty,$$

where $g_u(x, y) := \mu(x)p_u(x, y) - \mu(x)\mu(y)$. In order to show these conditions, an upper bound of the second derivative of the transition density $p_t(x, y)$ is obtained (see Lemma 1 below), for which the additional condition **A6** is needed.

As shown in [13], conditions **WLC1** and **WLC2** are also useful to show the asymptotic normality of the kernel density estimator, as proved in the next theorem.

Theorem 2. *Let X be the solution to* (2) *on* $[0, T]$ *with* $d = 1$. *Suppose that Assumptions* **A1**-**A6** *hold and* $\mu \in \mathcal{H}_1(\beta, \mathcal{L})$. *Then, for any collection* $(x_i)_{1\leq i\leq m}$ *of distinct real numbers*

$$\sqrt{T}(\hat{\mu}_{h,T}(x_i) - \mathbb{E}[\hat{\mu}_{h,T}(x_i)],\, 1 \leq i \leq m) \xrightarrow{\mathcal{D}} N^{(m)}(0, \Sigma^{(m)}) \text{ as } T \to \infty, \tag{6}$$

where

$$\Sigma^{(m)} := (\sigma(x_i, x_j))_{1\leq i,j\leq m}, \qquad \sigma(x_i, x_j) := 2\int_0^\infty g_u(x_i, x_j)du.$$

Furthermore,

$$\sqrt{T}(\hat{\mu}_{h,T}(x_i) - \mu(x_i),\, 1 \leq i \leq m) \xrightarrow{\mathcal{D}} N^{(m)}(0, \Sigma^{(m)}) \text{ as } T \to \infty. \tag{7}$$

We are also interested in obtaining lower bounds in dimension $d \in \{1, 2\}$. For this, we consider the particular case of equation (2):

$$X_t = X_0 + \int_0^t b(X_s)ds + aB_t + \int_0^t \int_{\mathbb{R}_0^d} \gamma z(\nu(ds, dz) - F(z)dzds), \tag{8}$$

where a and γ are $d \times d$ invertible matrices and b is a Lipschitz and bounded function satisfying Assumption **A2**. We assume that F satisfies Assumptions **A3**-**A5** and $\int_{\mathbb{R}^d} F(z)dz < \infty$. Then, the unique solution to equation (8) admits a unique invariant measure π_b, which we assume has a density μ_b with respect to the Lebesgue measure. We denote by $\mathbb{P}_b^{(T)}$ and $\mathbb{E}_b^{(T)}$ the law and expectation of the solution $(X_t)_{t\in[0,T]}$.

We say that a bounded and Lipschitz function b belongs to $\Sigma(\beta, \mathcal{L})$ if the unique invariant density μ_b belongs to $\mathcal{H}_d(\beta, \mathcal{L})$ for some $\beta, \mathcal{L} \in \mathbb{R}^d$, $\beta_i > 1$, $\mathcal{L}_i > 0$.

We define the minimax risk at a point $x \in \mathbb{R}^d$ by

$$\mathcal{R}_T^x(\beta, \mathcal{L}) := \inf_{\tilde{\mu}_T} \mathcal{R}(\tilde{\mu}_T(x)) := \inf_{\tilde{\mu}_T} \sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b^{(T)}[(\tilde{\mu}_T(x) - \mu_b(x))^2], \tag{9}$$

where the infimum is taken on all possible estimators of the invariant density.

The following lower bounds hold true.

Theorem 3. *Let X be the solution to* (8) *on* $[0, T]$ *with* $d = 1$. *We assume that* a *and* γ *are non-zero constants. There exists* $T_0 > 0$ *and* $c > 0$ *such that, for all* $T \geq T_0$,

$$\inf_{x\in\mathbb{R}} \mathcal{R}_T^x(\beta, \mathcal{L}) \geq \frac{c}{T}.$$

Theorem 4. *Let X be the solution to* (8) *on* $[0, T]$ *with* $d = 2$. *Assume that for all* $i \in \{1, 2\}$ *and* $j \neq i$,

$$|(aa^T)_{ij}(aa^T)_{jj}^{-1}| \leq \frac{1}{2}. \tag{10}$$

There exists $T_0 > 0$ *and* $c > 0$ *such that, for* $T \geq T_0$,

$$\inf_{\tilde{\mu}_T} \sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b^{(T)} \left[\sup_{x\in\mathbb{R}^2} (\tilde{\mu}_T(x) - \mu_b(x))^2 \right] \geq c\frac{\log T}{T}.$$

Comparing these lower bounds with the upper bound of Theorem 1 for the case $d = 1$ and Proposition 4 in [2] for the two-dimensional case, we conclude that the convergence rate $\{\frac{1}{T}, \frac{\log T}{T}\}$ are the best possible for the kernel estimator of the invariant density in dimension $d \in \{1, 2\}$.

The proof of Theorem 3 follows along the same lines as that of Theorem 2 in [3], where a lower bound for the kernel estimator of the invariant density for the solution to (8) for $d \geq 3$ is obtained. The proof is based on the two hypotheses method, explained for example in Section 2.3 of [37]. However, this method does not work for the two-dimensional case as explained in Remark 1 below. Instead, we use the Kullback's version of the finite number of hypotheses method as stated in Lemma C.1 of [36], see Lemma 2 below. Observe that this method gives a slightly weaker lower bound as we get a $\sup_x$ inside the expectation, while the method in [3] provides an $\inf_x$ outside the expectation.

3 Preliminary results

The proof of Theorems 1 and 2 will use the following upper bound on the second partial derivative of the transition density.

Lemma 1. *Let X be the solution to* (2) *on* $[0,T]$ *with* $d=1$. *Suppose that Assumptions* **A1-A6** *hold. For all* $T>0$, *there exist two constants* $\lambda_1>0$ *and* $c>0$ *such that for any* $x,y\in\mathbb{R}$ *and* $t\in[0,T]$

$$\frac{\partial^2}{\partial x^2}p_t(x,y)\le c\left(t^{-3/2}e^{-\lambda_1\frac{|y-x|^2}{t}}+\frac{1}{(t^{1/2}+|x-y|)^{1+\alpha}}\right).$$

Proof. We apply the estimate in Theorem 3.5(v) of [16]. We remark that, in [16], the authors assumed $d\ge 2$. After inspection of the proof it is possible to see that the result can be extended to the case $d=1$: it was stated for $d\ge 2$ for the convenience in describing the Kato class function (for the drift). We also remark that the sufficient conditions Theorem 3.5(v) of [16] are the same as that to obtain the upper bound for the transition density (3) (which hold under Assumptions **A1-5**), together with the following additional condition: there exist $c>0$ and $\delta\in(0,1)$ such that for all $x,y,z\in\mathbb{R}$,

$$|b(x)-b(y)|+|k(x,z)-k(y,z)|\le c|x-y|^\delta, \tag{11}$$

where $k(x,z)=\frac{1}{\gamma(x)}|z|^{1+\alpha}F(\frac{z}{\gamma(x)})$. Thus, we only need to show (11). As b is bounded and Lipschitz, it satisfies (11). In fact, when x and y are such that $|x-y|>1$, thanks to the boundedness of b we have, for each $\delta\in(0,1)$,

$$|b(x)-b(y)|\le|b(x)|+|b(y)|\le 2c\le 2c|x-y|^\delta.$$

Instead, when x and y are such that $|x-y|\le 1$, the Lipschitz continuity gives

$$|b(x)-b(y)|\le L|x-y|=L|x-y|^{1-\delta}|x-y|^\delta\le L|x-y|^\delta.$$

Concerning k, we write

$$\begin{aligned}|k(x,z)-k(y,z)|&=|z|^{1+\alpha}\left|\frac{1}{\gamma(x)}F\left(\frac{z}{\gamma(x)}\right)-\frac{1}{\gamma(y)}F\left(\frac{z}{\gamma(y)}\right)\right|\\&=\frac{|z|^{1+\alpha}}{|\gamma(x)|}\left|F\left(\frac{z}{\gamma(x)}\right)-F\left(\frac{z}{\gamma(y)}\right)\right|+|z|^{1+\alpha}\left|F\left(\frac{z}{\gamma(y)}\right)\right|\left|\frac{1}{\gamma(x)}-\frac{1}{\gamma(y)}\right|.\end{aligned} \tag{12}$$

From the intermediate value theorem and defining $\tilde z\in[\frac{z}{\gamma(x)},\frac{z}{\gamma(y)}]$ (assuming WLOG that $\gamma(x)>\gamma(y)$, otherwise $\tilde z\in[\frac{z}{\gamma(y)},\frac{z}{\gamma(x)}]$), the first term in the r.h.s above is bounded by

$$\begin{aligned}\frac{|z|^{1+\alpha}}{|\gamma(x)|}|F'(\tilde z)|\left|\frac{z}{\gamma(x)}-\frac{z}{\gamma(y)}\right|&\le\frac{|z|^{1+\alpha}}{|\gamma(x)|}\frac{c}{|\tilde z|^{2+\alpha}}\frac{|z|}{|\gamma(x)\gamma(y)|}|\gamma(y)-\gamma(x)|\\&\le c\frac{\gamma_{max}^{2+\alpha}}{\gamma_{min}^3}|\gamma(x)\gamma(y)|,\end{aligned}$$

where we have used **A6** in the first inequality and $\gamma_{\max}:=\sup_x|\gamma(x)|$ and $\gamma_{\min}:=\inf_x|\gamma(x)|$. Moreover, by **A3**, the second term in the r.h.s of (12) is bounded by

$$|z|^{1+\alpha}\frac{c|\gamma(y)|^{1+\alpha}}{|z|^{1+\alpha}}\frac{1}{|\gamma(x)\gamma(y)|}|\gamma(y)-\gamma(x)|\le c\frac{\gamma_{max}^\alpha}{\gamma_{min}}|\gamma(y)-\gamma(x)|.$$

Thus, we have shown that

$$|k(x,z)-k(y,z)| \le c\left(\frac{\gamma_{max}^{2+\alpha}}{\gamma_{min}^{3}}+\frac{\gamma_{max}^{\alpha}}{\gamma_{min}}\right)|\gamma(y)-\gamma(x)|.$$

Finally, as γ is Lipschitz and bounded, we conclude that (11) holds. This concludes the proof of the lemma. □

The key point of the proof of Theorem 1 consists in showing that conditions **WCL1** and **WCL2** hold true, which is proved in the next proposition.

Proposition 1. *Let X be the solution to* (2) *on $[0,T]$ with $d=1$. Suppose that Assumptions* **A1-A6** *hold. Then, conditions* **WCL1** *and* **WCL2** *are satisfied.*

Proof. We start considering **WCL1**. The density estimate (3) yields

$$p_t(x,y) \le ct^{-\frac{1}{2}} + \tilde{c}t^{\frac{1-\alpha}{2}} \le \bar{c}t^{-\frac{1}{2}} \qquad 0 < t \le 1, \tag{13}$$

which combined with $\sup_{y\in\mathbb{R}}\mu(y) < \infty$ gives **WCL1**. In order to show **WCL2**, we set $\varphi(\lambda) := \mathbb{E}[\exp(i\lambda X_t)]$ and $\varphi_x(\lambda,t) := \mathbb{E}[\exp(i\lambda X_t)|X_0 = x]$ and we claim that here exists $c_1 > 0$ such that

$$|\varphi(\lambda)| \le c_1(1+|\lambda|^{-2}). \tag{14}$$

Moreover, for all $t \ge 2$, there exists $c_2 > 0$, such that for all $x \in \mathbb{R}$,

$$|\varphi_x(\lambda,t)| \le c_2(1+|\lambda|^{-2}). \tag{15}$$

Recall from Lemma 2 in [2] that the process X is exponentially β-mixing, which implies that $\beta_X(u) \le ce^{-\gamma_1 u}$, where $\beta_X(u)$ is the β-mixing coefficient defined in Section 1.3.2 of [26]. It follows that, for any $p > 0$, $\int_0^\infty \beta_X^p(u)du < \infty$. Thus, by Proposition 10 of [18], inequalities (14) and (15) and the integrability of the β-mixing coefficient imply **WCL2**. Therefore, we are left to show (14) and (15). We start showing (15). Integrating by parts and using Lemma 1 it yields

$$\begin{aligned}
|\varphi_x(\lambda,t)| &= \left|\int_{\mathbb{R}} \exp(i\lambda y)p_t(x,y)dy\right| = |\lambda|^{-2}\left|\int_{\mathbb{R}} \exp(i\lambda y)\frac{\partial^2}{\partial y^2}p_t(x,y)dy\right| \\
&= |\lambda|^{-2}\left|\int_{\mathbb{R}} \exp(i\lambda y)\left(\frac{\partial^2}{\partial y^2}\int_{\mathbb{R}} p_{t-1}(x,z)p_1(z,y)\right)dy\right| \\
&\le |\lambda|^{-2}\int_{\mathbb{R}}\int_{\mathbb{R}} p_{t-1}(x,z)\left|\frac{\partial^2}{\partial y^2}p_1(z,y)\right|dz\,dy \\
&\le c|\lambda|^{-2}\int_{\mathbb{R}}\int_{\mathbb{R}} p_{t-1}(x,z)\left(e^{-\frac{\lambda_1}{2}|x-y|^2} + \frac{1}{(1+|x-y|)^{1+\alpha}}\right)dy\,dz.
\end{aligned}$$

As $\alpha \in (0,2)$, the integral in dy is finite. Since $\int_{\mathbb{R}} p_{t-1}(x,z)dz < c$, we get

$$|\varphi_x(\lambda,t)| \le c|\lambda|^{-2} \le c(1+|\lambda|^{-2}),$$

which proves (15). Similarly,

$$\begin{aligned}
|\varphi(\lambda)| &= \left|\int_{\mathbb{R}} \exp(i\lambda y)\mu(y)dy\right| = |\lambda|^{-2}\left|\int_{\mathbb{R}} \exp(i\lambda y)\left(\frac{\partial^2}{\partial y^2}\int_{\mathbb{R}} \mu(z)p_1(z,y)\right)dy\right| \\
&\le |\lambda|^{-2}\int_{\mathbb{R}}\int_{\mathbb{R}} \mu(z)\left|\frac{\partial^2}{\partial y^2}p_1(z,y)\right|dz\,dy \le c(1+|\lambda|^{-2}),
\end{aligned}$$

which gives (15). The proof of the proposition is now completed. □

Theorem 2 is an application of the following central limit theorem for discrete stationary sequences. Let $Y_n = (Y_{n,i},\, i \in \mathbb{Z})$, $n \geq 1$ be a sequence of strictly stationary discrete time $\mathbb{R}^m$ valued random process. We define the α-mixing coefficient of Y_n by

$$\alpha_{n,k} := \sup_{A \in \sigma(Y_{n,i},\, i \leq 0),\quad B \in \sigma(Y_{n,i},\, i \geq k)} \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$$

and we set $\alpha_k := \sup_{n \geq 1} \alpha_{n,k}$ (see also Section 1 in [26]). We denote by $Y^{(r)}$ the r-th component of an m dimensional random vector Y.

Theorem 5 (Theorem 1.1 [13])**.** *Assume that*

(i) $\mathbb{E}[Y_{n,i}^{(r)}] = 0$ *and* $|Y_{n,i}^{(r)}| \leq M_n$ *for every* $n \geq 1$, $i \geq 1$ *and* $1 \leq r \leq m$, *where* M_n *is a constant depending only* n.

(ii)

$$\sup_{i \geq 1, 1 \leq r \leq m} \mathbb{E}[(Y_{n,i}^{(r)})^2] < \infty.$$

(iii) *For every* $1 \leq r, s \leq m$ *and for every sequence* $b_n \to \infty$ *such that* $b_n \leq n$ *for every* $n \geq 1$, *we have*

$$\lim_{n \to \infty} \frac{1}{b_n} \mathbb{E}\left[\sum_{i=1}^{b_n} Y_{n,i}^{(r)} \sum_{j=1}^{b_n} Y_{n,j}^{(s)}\right] = \sigma_{r,s}.$$

(iv) *There exists* $a \in (1, \infty)$ *such that* $\sum_{k \geq 1} k \alpha_k^{\frac{a-1}{a}} < \infty$.

(v) *For some constant* $c > 0$ *and for every* $n \geq 1$, $M_n \leq c n^{\frac{a^2}{(3a-1)(2a-1)}}$.

Then,

$$\frac{\sum_{i=1}^n Y_{n,i}}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \Sigma) \quad \text{as } n \to \infty,$$

where $\Sigma = (\sigma_{r,s})_{1 \leq r,s \leq m}$.

The proof of Theorem 4 is based on the following Kullback version of the main theorem on lower bounds in [37], see Lemma C.1 of [36]:

Lemma 2. *Fix* $\beta, \mathcal{L} \in (0,\infty)^2$ *and assume that there exists* $f_0 \in \mathcal{H}_2(\beta, \mathcal{L})$ *and a finite set* J_T *such that one can find* $\{f_j,\, j \in J_T\} \subset \mathcal{H}_2(\beta, \mathcal{L})$ *satisfying*

$$\|f_j - f_k\|_\infty \geq 2\psi > 0 \qquad \forall j \neq k \in J_T. \tag{16}$$

Moreover, denoting $\mathbb{P}_j$ *the probability measure associated with* f_j, $\forall j \in J_T$, $\mathbb{P}_j^{(T)} \ll \mathbb{P}_0^{(T)}$ *and*

$$\frac{1}{|J_T|} \sum_{j \in J_T} KL(\mathbb{P}_j^{(T)}, \mathbb{P}_0^{(T)}) = \frac{1}{|J_T|} \sum_{j \in J_T} \mathbb{E}_j^{(T)}\left[\log\left(\frac{d\mathbb{P}_j^{(T)}}{d\mathbb{P}_0^{(T)}}(X^T)\right)\right] \leq \gamma \log(|J_T|) \tag{17}$$

for some $\gamma \in (0, \frac{1}{8})$. *Then, for* $q > 0$, *we have*

$$\inf_{\tilde{\mu}_T} \sup_{\mu_b \in \mathcal{H}_2(\beta, \mathcal{L})} (\mathbb{E}_b^{(T)}[\psi^{-q} \|\tilde{\mu}_T - \mu_b\|_\infty^q])^{1/q} \geq c(\gamma) > 0,$$

where the infimum is taken over all the possible estimators $\tilde{\mu}_T$ *of* μ_b.

4 Proof of the main results

4.1 Proof of Theorem 1

By the symmetry of the covariance operator and the stationarity of the process,

$$\begin{aligned}
T\,\mathrm{Var}(\hat{\mu}_{h,T}(x)) &= \frac{1}{T}\int_0^T\int_0^T \mathrm{Cov}(\mathbb{K}_h(x-X_t),\mathbb{K}_h(x-X_s))ds\,dt \\
&= \frac{2}{T}\int_0^T (T-u)\mathrm{Cov}(\mathbb{K}_h(x-X_u),\mathbb{K}_h(x-X_0))du \\
&= 2\int_0^T (1-\frac{u}{T})\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{K}_h(x-y)\mathbb{K}_h(x-z)g_u(y,z)dy\,dz\,du \\
&\leq \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{K}_h(x-y)\mathbb{K}_h(x-z)\int_0^\infty g_u(y,z)dy\,dz\,du \leq c,
\end{aligned}$$

where in the last inequality we have used Proposition 1. Then, from the bias-variance decomposition (5) we obtain (4), which concludes the desired proof.

4.2 Proof of Theorem 2

We aim to apply Theorem 5. First of all we split the interval $[0,T]$ into n small intervals whose length is Δ_n as follows: $[0,T] = \cup_{i=1}^n [t_{i-1},t_i)$, with $t_0 = 0$ and $t_n = T$ and, for any $i \in \{1,\dots,n\}$, $t_i - t_{i-1} = \Delta_n$. By construction, it clearly holds that $n\Delta_n = T$.

For each $n \geq 1$ and $1 \leq r \leq m$, we consider the sequence $(Y_{n,i}^{(r)})_{i\geq 1}$ defined as

$$Y_{n,i}^{(r)} := \frac{1}{\sqrt{\Delta_n}}\left(\int_{t_{i-1}}^{t_i} \mathbb{K}_h(x_r - X_u)du - \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \mathbb{K}_h(x_r - X_u)du\right]\right),$$

for $x_r \in \mathbb{R}$. We denote by $Y_{n,i}$ the $\mathbb{R}^m$ valued random vector defined by $Y_{n,i} = (Y_{n,i}^{(1)},\dots,Y_{n,i}^{(m)})$. By construction,

$$\frac{\sum_{i=1}^n Y_{n,i}}{\sqrt{n}} = \sqrt{T}(\hat{\mu}_{h,T}(x) - \mathbb{E}[\hat{\mu}_{h,T}(x)]),$$

where $\hat{\mu}_{h,T}(x) - \mathbb{E}[\hat{\mu}_{h,T}(x)]$ is the vector

$$(\hat{\mu}_{h,T}(x_1) - \mathbb{E}[\hat{\mu}_{h,T}(x_1)],\dots,\hat{\mu}_{h,T}(x_m) - \mathbb{E}[\hat{\mu}_{h,T}(x_m)]).$$

It is clear that $\mathbb{E}[Y_{n,i}] = 0$ for all $n \geq 1$ and $i \geq 1$. Moreover, for all $i \geq 1$, $1 \leq r \leq m$ and $n \geq 1$ we have

$$|Y_{n,i}^{(r)}| \leq \frac{1}{\sqrt{\Delta_n}}\left\|\mathbb{K}_h\right\|_\infty \Delta_n \leq \frac{c}{h(T)}\sqrt{\Delta_n} =: M_n.$$

Hence, assumption (i) holds true. Concerning assumption (ii) we remark that, for any $i \geq 1$ and any $1 \leq r \leq m$,

$$\begin{aligned}
\mathbb{E}[(Y_{n,i}^{(r)})^2] &= \mathrm{Var}\left(\frac{1}{\sqrt{\Delta_n}}\int_0^{\Delta_n} \mathbb{K}_h(x_r - X_u)du\right) = \mathrm{Var}(\sqrt{\Delta_n}\hat{\mu}_{h,\Delta_n}(x_r)) \\
&= \Delta_n \mathrm{Var}(\hat{\mu}_{h,\Delta_n}(x_r)) \leq \Delta_n\frac{c}{\Delta_n} = c,
\end{aligned}$$

where in the last inequality we have used (4.1). We next check condition (iii). Let b_n be a sequence of integers such that $b_n \to \infty$ and $b_n \le n$ for every n. For every $1 \le r \le m$ and $1 \le s \le m$, we have

$$\frac{1}{b_n}\mathbb{E}\left[\sum_{i=1}^{b_n} Y_{n,i}^{(r)} \sum_{j=1}^{b_n} Y_{n,j}^{(s)}\right] = \frac{1}{b_n}\int_0^{b_n}\int_0^{b_n} \mathrm{Cov}(\mathbb{K}_h(x_r - X_u), \mathbb{K}_h(x_s - X_v))du\, dv$$
$$= 2\int_0^{b_n}(1-\frac{u}{b_n})\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{K}_h(x_r - z_1)\mathbb{K}_h(x_s - z_2)g_u(z_1, z_2)dz_1\, dz_2\, du$$
$$= 2\int_{\mathbb{R}}\int_{\mathbb{R}}\int_0^{b_n}(1-\frac{u}{b_n})K(w_1)K(w_2)g_u(x_r - h(T)w_1, x_s - h(T)w_2)du\, dw_1\, dw_2,$$

where we have used Fubini's theorem and the change of variables $w_1 := \frac{x_r - z_1}{h(T)}$, $w_2 := \frac{x_s - z_2}{h(T)}$. Using dominated convergence and the fact that $h(T) \to 0$ for $T \to \infty$, we obtain

$$\lim_{n\to\infty}\frac{1}{b_n}\mathbb{E}\left[\sum_{i=1}^{b_n} Y_{n,i}^{(r)} \sum_{j=1}^{b_n} Y_{n,j}^{(s)}\right] = 2\int_{\mathbb{R}} K(w_1)\int_{\mathbb{R}} K(w_2)\int_0^{\infty} g_u(x_r, x_s)du\, dw_2\, dw_1$$
$$= 2\int_0^{\infty} g_u(x_r, x_s)du =: \sigma(x_r, x_s),$$

which proves (iii). Remark that it possible to use dominated convergence theorem as we have shown in Proposition 1 that $\int_0^\infty \|g_u\|_\infty du < \infty$. In particular, we have

$$|(1-\frac{u}{b_n})K(w_1)K(w_2)g_u(x_r - h(T)w_1, x_s - h(T)w_2)1_{[0,b_n]}(u)1_{\mathbb{R}^2}(w_1, w_2)|$$
$$\le \|g_u\|_\infty |K(w_1)K(w_2)| \in L^1(\mathbb{R}^+ \times \mathbb{R}^2).$$

We now check (iv). We remark that if a process is β-mixing, then it is also α-mixing and the following estimation holds (see Theorem 3 in Section 1.2.2 of [26])

$$\alpha_k \le \beta_{Y_{n,i}}(k) = \beta_X(k) \le ce^{-\gamma_1 k}.$$

Therefore, we it suffices to show that there exists $a \in (1, \infty)$ such that

$$c\sum_{k\ge 1} ke^{-k\gamma_1\frac{(a-1)}{a}} < \infty,$$

which is true for any $a > 1$, so (iv) is satisfied.

We are left to show (v). Set $f(a) := \frac{a^2}{(3a-1)(2a-1)}$. We want to show that there exists $a > 1$ such that, for some $c > 0$ and for any $n \ge 1$,

$$\frac{\sqrt{\Delta_n}}{h(T)} \le cn^{f(a)}. \tag{18}$$

For any $\epsilon > 0$, we can choose $h(T) = (\frac{1}{T})^{\frac{1}{2\beta}+\epsilon}$ which still achieves the rate optimal choice of Theorem 1. Recalling that $T = n\Delta_n$ and replacing it in (18), we get

$$c\sqrt{\Delta_n}(n\Delta_n)^{\frac{1}{2\beta}+\epsilon} \le n^{f(a)},$$

which holds true if and only if

$$\Delta_n^{\frac{\beta+1+\epsilon 2\beta}{2\beta}} \leq cn^{\frac{2\beta f(a)-1-\epsilon 2\beta}{2\beta}}.$$

This condition is satisfied if for some $\epsilon' > 0$,

$$\Delta_n \leq cn^{\frac{2\beta f(a)-1}{\beta+1}-\epsilon'}.$$

We can always find an $\epsilon' > 0$ such that this last condition holds true. In fact, it is easy to see that for any $\beta > 1$ there exists $a > 1$ such that $f(a) > \frac{1}{2\beta}$ so that $\frac{2\beta f(a)-1}{\beta+1} > 0$. Thus, condition (v) is satisfied. We can then apply Theorem 5 which directly leads us to (6). We next turn to the proof of (7). In the proof of Theorem 1 we have shown that

$$|\mathbb{E}[\hat{\mu}_{h,T}(x_i)] - \mu(x_i)| \leq h(T)^\beta.$$

where $h(T) = (\frac{1}{T})^{\bar{a}}$, with $\bar{a} > \frac{1}{2\beta}$. Thus,

$$\sqrt{T}|\mathbb{E}[\hat{\mu}_{h,T}(x_i)] - \mu(x_i)| \leq cT^{\frac{1}{2}}T^{-\bar{a}\beta},$$

which converges to zero. This proves (7) and concludes the desired proof.

4.3 Proof of Theorem 3

The proof of of Theorem 3 follows as the proof of the lower bound for $d \geq 3$ obtained in Theorem 3 of [3]. Therefore, we will only explain the main steps and the principal differences.

Step 1 The first step consists in showing that given a density function f, we can always find a drift function b_f such that f is the unique invariant density function of equation (8) with drift coefficient $b = b_f$. We give the statement and proof in dimension $d = 1$, as in Propositions 2 and 3 of [3] it is only done for $d \geq 2$.

Proposition 2. *Let $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^2$ positive probability density satisfying the following conditions*

1. $\lim_{y\to\pm\infty} f(y) = 0$ *and* $\lim_{y\to\pm\infty} f'(y) = 0$.
2. *There exists $0 < \epsilon < \frac{\epsilon_0}{|\gamma|}$, where ϵ_0 is as in Assumption* **A4** *such that, for any $y, z \in \mathbb{R}$,*
$$f(y \pm z) \leq \hat{c}_1 e^{\epsilon|z|} f(y).$$
3. *For $\epsilon > 0$ as in 2. there exists $\hat{c}_2(\epsilon) > 0$ such that*
$$\sup_{y<0} \frac{1}{f(y)} \int_{-\infty}^{y} f(w)dw < \hat{c}_2 \quad \text{and} \quad \sup_{y>0} \frac{1}{f(y)} \int_{y}^{\infty} f(w)dw < \hat{c}_2.$$
4. *There exists $0 < \tilde{\epsilon} < \frac{a^2}{2\gamma^2 c_4 \hat{c}_2 \hat{c}_4 \hat{c}_1}$ and $R > 0$ such that for any $|y| > R$, $\frac{f'(y)}{f(y)} \leq -\tilde{\epsilon}\,\mathrm{sgn}(y)$, where c_4 is as in Assumption* **A4**. *Moreover, there exists $\hat{c}_3$ such that for any $y \in \mathbb{R}$, $|f'(y)| \leq \hat{c}_3 f(y)$.*

5. For any $y \in \mathbb{R}$ and $\tilde{\epsilon}$ as in 4. $|f''(y)| \leq \hat{c}_4 \tilde{\epsilon}^2 f(y)$.

Then there exists a bounded Lipschitz function b_f which satisfies **A2** *such that f is the unique invariant density to equation* (8) *with drift coefficient $b = b_f$.*

Proof. Let A_d be the discrete part of the generator of the diffusion process X solution of (8) and let A_d^* its adjoint. We define b_f as

$$b_f(x) = \begin{cases} \frac{1}{f(x)} \int_{-\infty}^{x} (\frac{1}{2} a^2 f''(w) + A_d^* f(w)) dw, & \text{if } x < 0; \\ -\frac{1}{f(x)} \int_{x}^{\infty} \frac{1}{2} a^2 f''(x)(w) + A_d^* f(w) dw, & \text{if } x > 0, \end{cases}$$

where

$$A_d^* f(x) = \int_{\mathbb{R}} [f(x - \gamma z) - f(x) + \gamma z f'(x)] F(z) dz.$$

Then, following Proposition 3 in [3], one can check that b_f is bounded, Lipschitz, and satisfies **A2**. Moreover, if we replace b by b_f in equation (8), then f is the unique invariant density. $\square$

Step 2 The second step consists in defining two probability density functions f_0 and f_1 in $\mathcal{H}_1(\beta, \mathcal{L})$.

We first define $f_0(y) = c_\eta f(\eta |y|)$, where $\eta \in (0, \frac{1}{2})$, c_η is such that $\int f_0 = 1$, and

$$f(x) = \begin{cases} e^{-|x|}, & \text{if } |x| \geq 1 \\ \in [1, e^{-1}], & \text{if } \frac{1}{2} < |x| < 1 \\ 1, & \text{if } |x| \leq \frac{1}{2}. \end{cases} \tag{19}$$

Moreover, f is a C^2 function such that for any $x \in \mathbb{R}$,

$$\frac{1}{2} e^{-|x|} \leq f(x) \leq 2e^{-|x|}, \quad |f'(|x|)| \leq 2e^{-|x|}, \quad \text{and} \quad |f'(|x|)| \leq 2e^{-|x|}.$$

It is easy to see that η can be chosen small enough so that $f_0 \in \mathcal{H}_1(\beta, \mathcal{L})$. Moreover, f_0 satisfies the assumptions of Proposition 2 with $\hat{c}_1 = 4$, $\epsilon = \eta$, $\hat{c}_2 = \frac{4}{\eta}$, $R = \frac{1}{\eta}$, $\hat{c}_3 = 4$, and $\hat{c}_4 = 16$. In order for the condition on $\tilde{\epsilon}$ in assumption 4. to be satisfied we need $c_4 < \frac{a^2}{2\gamma^2 4^4}$. This means that the jumps have to integrate an exponential function. The bound depends on the coefficients a and γ and so it depends only on the model.

Therefore, $b_0 := b_{f_0}$ belongs to $\Sigma(\beta, \mathcal{L})$. Recall that b_0 belongs to $\Sigma(\beta, \mathcal{L})$ if and only if f_0 belongs to $\mathcal{H}_1(\beta, \mathcal{L})$ and b_0 is bounded, Lipschitz and satisfies the drift condition **A2**.

We next define

$$f_1(x) = f_0(x) + \frac{1}{M_T} \hat{K} \left(\frac{x - x_0}{H(T)} \right), \tag{20}$$

where $x_0 \in \mathbb{R}$ is fixed and $\hat{K} : \mathbb{R} \to \mathbb{R}$ is a C^∞ function with support on $[-1, 1]$ such that

$$\hat{K}(0) = 1, \quad \int_{-1}^{1} \hat{K}(z) dz = 0.$$

M_T and $H(T)$ will be calibrated later and satisfy that $M_T \to \infty$ and $H(T) \to 0$, as $T \to \infty$.

Then it can be shown as in [3, Lemma 3] that if there exists $\epsilon > 0$ small enough such that for all T sufficiently large,

$$\frac{1}{M_T} \leq \epsilon H(T)^\beta \qquad \text{and} \qquad \frac{1}{H(T)} = o(M_T) \tag{21}$$

as $T \to \infty$, then $b_1 := b_{f_1}$ belongs to $\Sigma(\beta, \mathcal{L})$ for T sufficiently large.

Step 3 As $b_0, b_1 \in \Sigma(\beta, \mathcal{L})$, we can write

$$R(\tilde{\mu}_T(x_0)) \geq \frac{1}{2}\mathbb{E}_1^{(T)}[(\tilde{\mu}_T(x_0) - f_1(x_0))^2] + \frac{1}{2}\mathbb{E}_0^{(T)}[(\tilde{\mu}_T(x_0) - f_0(x_0))^2],$$

where $\mathbb{E}_i^{(T)}$ denotes the expectation with respect to b_i. Then, following as in [3], using Girsanov's formula, we can show that if

$$\sup_{T \geq 0} T \frac{1}{M_T^2 H(T)} < \infty, \tag{22}$$

then for sufficiently large T,

$$R(\tilde{\mu}_T(x_0)) \geq \frac{C}{8\lambda} \frac{1}{M_T^2}, \tag{23}$$

where the constants C and λ are as in Lemma 4 of [3] and they do not depend on the point x_0. We finally look for the larger choice of $\frac{1}{M_T^2}$ for which both (21) and (22) hold true. It suffices to choose $M_T = \sqrt{T}$ and $H(T)$ a constant, to conclude the proof of Theorem 3.

Remark 1. *The two hypothesis method used above does not work to prove the 2-dimensional lower bound of Theorem 4. Indeed, following as above, we can define*

$$f_1(x) = f_0(x) + \frac{1}{M_T} \hat{K}\left(\frac{x - x_0}{H_1(T)}\right) \hat{K}\left(\frac{x - x_0}{H_2(T)}\right).$$

Then, it is possible to show that (23) *still holds and, therefore, we should take* M_T *such that* $\frac{1}{M_T^2} = \frac{\log T}{T}$. *On the other hand, condition* (22) *now becomes*

$$\sup_{T \geq 0} T \frac{1}{M_T^2} \left(\frac{H_2(T)}{H_1(T)} + \frac{H_1(T)}{H_2(T)}\right) < \infty.$$

The optimal choice of the bandwidth is achieved for $H_2(T) = H_1(T)$ *which yields to* $\sup_{T \geq 0} T \frac{1}{M_T^2} < \infty$, *which is clearly not satisfied when* $\frac{1}{M_T^2} = \frac{\log T}{T}$.

4.4 Proof of Theorem 4

We will apply Lemma 2 with $\psi := v\sqrt{\frac{\log T}{T}}$, where $v > 0$ is fixed. As above we divide the proof into three steps.

Step 1 As in the one-dimensional case, the first step consists in showing that given a density function f, we can always find a drift function b_f such that f is the unique invariant density function of equation (8) with drift coefficient $b = b_f$, which is proved in Propositions 2 and 3 of [3]. We remark that condition (10) is needed in Proposition 3 to ensure that the terms on the diagonal of the volatility coefficient a dominate on the others, which is crucial to get that b_f satisfies the drift condition **A2**.

Step 2 We next define the probability density $f_0 \in \mathcal{H}_2(\beta, \mathcal{L})$, the finite set J_T, and the set of probability densities $\{f_j,\, j \in J_T\} \subset \mathcal{H}_2(\beta, \mathcal{L})$ needed in order to apply Lemma 2.

We first define f_0 as π_0 in Section 7.2 of [3], which is the two-dimensional version of f_0 defined in the proof of Theorem 3, that is,

$$f_0(x) = c_\eta f(\eta (aa^T)^{-1}_{11}|x_1|) f(\eta (aa^T)^{-1}_{22}|x_2|), \quad x = (x_1, x_2) \in \mathbb{R}^2, \tag{24}$$

where f is as in (19). The density f_0 belongs to $\mathcal{H}_2(\beta, \mathcal{L})$ by construction.

We then set

$$J_T := \left\{1, \dots, \lfloor \frac{1}{\sqrt{H_1}} \rfloor \right\} \times \left\{1, \dots, \lfloor \frac{1}{\sqrt{H_2}} \rfloor \right\}, \tag{25}$$

where $H_1 := H_1(T)$ and $H_2 := H_2(T)$ are two quantities that converge to 0 as $T \to \infty$ and need to be calibrated.

Finally, for $j := (j_1, j_2) \in J_T$, we define $x_j := (x_{j,1}, x_{j,2}) = (j_1 H_1, j_2 H_2)$ and we set

$$f_j(x) := f_0(x) + v \sqrt{\frac{\log T}{T}} \hat{K} \left(\frac{x_1 - x_{j,1}}{H_1} \right) \hat{K} \left(\frac{x_2 - x_{j,2}}{H_2} \right),$$

where recall that $v > 0$ is fixed and $\hat{K}$ is as in (20).

Acting as in Lemma 3 of [3], recalling that the rate $\frac{1}{M_T}$ therein is now replaced by $\sqrt{\frac{\log T}{T}}$ (see also points 1. and 3. in the proof of Proposition 3 below), it is easy to see that if there exists $\epsilon > 0$ sufficiently small such that for large T,

$$\sqrt{\frac{\log T}{T}} \le \epsilon H_1^{\beta_1}, \qquad \sqrt{\frac{\log T}{T}} \le \epsilon H_2^{\beta_2}, \tag{26}$$

then, for any $j \in J_T$ and large T, $b_j \in \Sigma(\beta, \mathcal{L})$. In particular, $f_j \in \mathcal{H}_2(\beta, \mathcal{L})$. Therefore, $\{f_j,\, j \in J_T\} \subset \mathcal{H}_2(\beta, \mathcal{L})$ and, by construction,

$$\|f_j - f_k\|_\infty \ge 2\, v \sqrt{\frac{\log T}{T}} \|\hat{K}\|_\infty^2 = 2\, v \sqrt{\frac{\log T}{T}},$$

which proves the first condition of Lemma 2.

Step 3 We are left to show the remaining conditions of Lemma 2. The absolute continuity $\mathbb{P}_j^{(T)} \ll \mathbb{P}_0^{(T)}$ and the expression for $\frac{d\mathbb{P}_j^{(T)}}{d\mathbb{P}_0^{(T)}}(X^T)$ are both obtained by Girsanov formula, as in Lemma 4 of [3]. We have,

$$KL(\mathbb{P}_j^{(T)}, \mathbb{P}_0^{(T)}) = \mathbb{E}_j^{(T)} \left[\log \left(\frac{f_j}{f_0}(X^T) \right) \right] + \frac{1}{2} \mathbb{E}_j^{(T)} \left[\int_0^T |a^{-1}(b_0(X_u) - b_j(X_u))|^2 du \right],$$

where the law of $X^T = (X_t)_{t \in [0,T]}$ under $\mathbb{P}_j^{(T)}$ is the one of the solution to equation (8) with $b = b_0$.

By the definition of the f_j's it is easy to see that the first term is $o(1)$ as $T \to \infty$. In fact, as $\hat{K}$ is supported in $[-1, 1]$,

$$\mathbb{E}_j^{(T)}\left[\log\left(\frac{f_j}{f_0}(X^T)\right)\right] = \int_{\mathbb{R}^2} \log\left(1 + \frac{v\sqrt{\frac{\log T}{T}}\hat{K}\left(\frac{x_1 - x_{j,1}}{H_1}\right)\hat{K}\left(\frac{x_2 - x_{j,2}}{H_2}\right)}{f_0(x)}\right) f_0(x)dx$$
$$\leq \left|\log\left(1 + c_* v\sqrt{\frac{\log T}{T}}\|\hat{K}\|_\infty^2\right)\right|,$$

which tends to zero as $T \to \infty$, where $c_* := \frac{4}{c_\eta}e^{4\eta k}$, c_η is the constant of normalization introduced in the definition of f_0, and $k := \max_{i=1,2}(aa^T)^{-1}_{ii}$. In fact, this follows from the definition of f_0 in (24). Since $f(x) \geq \frac{1}{2}e^{-|x|}$, we obtain

$$\frac{1}{f_0(x)} \leq \frac{1}{c_\eta}\frac{2}{e^{-\eta(aa^T)^{-1}_{11}|x_1|}}\frac{2}{e^{-\eta(aa^T)^{-1}_{22}|x_2|}} \leq \frac{4}{c_\eta}e^{\eta k(|H_1| + |x_{j,1}| + |H_2| + |x_{j,2}|)},$$

where we have also used the fact that, as $\hat{K}$ is supported in $[-1, 1]$, we have $x \in [x_{j,1} - H_1, x_{j,1} + H_1] \times [x_{j,2} - H_2, x_{j,2} + H_2]$. Finally, by the definition of x_j and the fact that $H_i \to 0$ as $T \to \infty$ for $i = 1, 2$ (and so for T large enough they are smaller than 1), we get

$$\frac{1}{f_0(x)} \leq \frac{4}{c_\eta}e^{4\eta k} \quad \text{for any } x \in [x_{j,1} - H_1, x_{j,1} + H_1] \times [x_{j,2} - H_2, x_{j,2} + H_2]. \tag{27}$$

Regarding the second term, using the stationarity of the process X^T, we have

$$\mathbb{E}_j^{(T)}\left[\int_0^T |a^{-1}(b_0(X_u) - b_j(X_u))|^2 du\right] = T\int_{\mathbb{R}^2} |a^{-1}(b_0(x) - b_j(x))|^2 f_0(x)dx.$$

Then, the following asymptotic bound will be proved at the end of this Section.

Proposition 3. *For T large enough,*

$$\int_{\mathbb{R}^2} |a^{-1}(b_0(x) - b_j(x))|^2 f_0(x)dx \leq 64\frac{e^{8\eta k}}{c_\eta^2}k^2 v^2 H_1 H_2\left(\frac{1}{H_1} + \frac{1}{H_2}\right)^2 \frac{\log T}{T}.$$

Taking the optimal choice for the bandwidth in Proposition 3, which is $H_1 = H_2$, we get that

$$\int_{\mathbb{R}^2} |a^{-1}(b_0(x) - b_j(x))|^2 f_0(x)dx \leq 64\frac{e^{8\eta k}}{c_\eta^2}k^2 v^2 4\frac{\log T}{T}.$$

In particular, after having ordered $\beta_1 \leq \beta_2$, we choose $H_1 = H_2 = (\frac{\log T}{T})^a$ with $a \leq \frac{1}{2\beta_2} = (\frac{1}{2\beta_1} \wedge \frac{1}{2\beta_2})$ so that condition (26) is satisfied. We therefore get

$$KL(\mathbb{P}_j^{(T)}, \mathbb{P}_0^{(T)}) \leq 128\frac{e^{8\eta k}}{c_\eta^2}k^2\, v^2 \log T \leq 128\frac{e^{8\eta k}}{c_\eta^2 a}k^2\, v^2 \log(|J_T|),$$

being the last estimation a consequence of the fact that, by construction,

$$\log(|J_T|) \geq a\log\left(\frac{T}{\log T}\right) = a\log(T)(1 + o(1)).$$

It is therefore enough to choose v such that $128\frac{e^{8\eta k}}{c_\eta^2 a}k^2\, v^2 < \frac{1}{8}$ (ie $v^2 < \frac{c_\eta^2 a}{1024\, k^2 e^{8\eta k}}$) and apply Lemma 2 to conclude the proof of Theorem 4.

4.5 Proof of Proposition 3

The proof of Proposition 3 follows similarly as Proposition 4 of [3]. Indeed, we first define the set

$$K_T^j := [x_{j,1} - H_1, x_{j,1} + H_1] \times [x_{j,2} - H_2, x_{j,2} + H_2]$$

and then show the following points for T large enough:

1. For any $x \in K_T^{j\,c}$ and $i \in \{1,2\}$: $|b_j^i(x) - b_0^i(x)| \leq c\, v \sqrt{\frac{\log T}{T}}$.
2. For any $i \in \{1,2\}$: $\int_{K_T^{j\,c}} |b_j^i(x) - b_0^i(x)| f_0(x) dx \leq c\, v \sqrt{\frac{\log T}{T}} H_1 H_2$.
3. For any $x \in K_T^j$ and $i \in \{1,2\}$: $|b_j^i(x) - b_0^i(x)| \leq \frac{8}{c_\eta} e^{4\eta k} k v \sqrt{\frac{\log T}{T}} \left(\frac{1}{H_1} + \frac{1}{H_2} \right)$.

The proof of the first two points follows exactly the one in Proposition 4 of [3], remarking that

$$d_T(x) := \pi_1(x) - \pi_0(x) = \frac{1}{M_T} \prod_{l=1}^{d} K\left(\frac{x_l - x_0^l}{h_l(T)} \right)$$

in [3] is now replaced by

$$d_T^j(x) := f_j(x) - f_0(x) = v \sqrt{\frac{\log T}{T}} \hat{K}\left(\frac{x_1 - x_{j,1}}{H_1} \right) \hat{K}\left(\frac{x_2 - x_{j,2}}{H_2} \right),$$

and the set

$$K_T := [x_0^1 - h_1(T), x_0^1 + h_1(T)] \times \cdots \times [x_0^d - h_d(T), x_0^d + h_d(T)]$$

introduced in the d-dimensional framework is now replaced by K_T^j. We recall that K and $\hat{K}$ are exactly the same kernel function. The proof of Proposition 4 of [3] is based on the fact that $d_T(x)$ and its derivatives are null for $x \in K_T^c$. In the same way, $d_T^j(x)$ and its derivatives are null for $x \in K_T^{j\,c}$. Then, acting as in [3], it is easy to see that the first two points above hold true.

Comparing the third point above with the third point of Proposition 4 of [3], it is clear that our goal is to make explicit the constant c. Keeping the notation in [3], we first introduce the following quantities:

$$\tilde{I}_1^i[f_0](x) := \frac{1}{2} \sum_{j=1}^{2} (aa^T)_{ij} \frac{\partial f_0}{\partial x_j}(x), \qquad \tilde{I}_2^i[f_0](x) = \int_{-\infty}^{x_i} A_{d,i}^* f_0(w_i) dw.$$

We moreover introduce the notation

$$\tilde{I}^i[f_0](x) = \tilde{I}_1^i[f_0](x) + \tilde{I}_2^i[f_0](x).$$

According with the definition of b, we have

$$b_0^i(x) = \frac{1}{f_0(x)} \tilde{I}^i[f_0](x), \qquad b_j^i(x) = \frac{1}{f_j(x)} \tilde{I}^i[f_j](x).$$

Since the operator $f \to \tilde{I}^i[f]$ is linear, we deduce that

$$b_j^i(x) = \frac{1}{f_j(x)}\tilde{I}^i[f_j](x) = \frac{1}{f_j(x)}\tilde{I}^i[f_0](x) + \frac{1}{f_j(x)}\tilde{I}^i[d_T^j](x). \tag{28}$$

Therefore,

$$b_j^i - b_0^i = (\frac{1}{f_j} - \frac{1}{f_0})\tilde{I}^i[f_0] + \frac{1}{f_j}\tilde{I}^i[d_T^j] = \frac{f_0 - f_j}{f_j}\frac{1}{f_0}\tilde{I}^i[f_0] + \frac{1}{f_j}\tilde{I}^i[d_T^j] = \frac{d_T^j}{f_j}b_0^i + \frac{1}{f_j}\tilde{I}^i[d_T^j].$$

We need to evaluate such a difference on the compact set K_T^j. For this, we will use that fact that $f_j = f_0 + d_T^j$, and obtain a lower bound away from 0. Specifically, from the definition of d_T^j, we get

$$\left\|d_T^j\right\|_\infty \leq v\sqrt{\frac{\log T}{T}}\|\hat{K}\|_\infty^2 = v\sqrt{\frac{\log T}{T}}. \tag{29}$$

In particular,

$$f_j \geq f_0 - |d_T^j| \geq f_0 - v\sqrt{\frac{\log T}{T}} \geq \frac{f_0}{2},$$

since $\sqrt{\frac{\log T}{T}} \to 0$ as $T \to \infty$, so for T large enough we have $v\sqrt{\frac{\log T}{T}} \leq \frac{f_0}{2}$. Then, for any $x \in K_T^j$, using (27) we have

$$\frac{1}{f_j(x)} \leq \frac{2}{f_0} \leq \frac{8}{c_\eta}e^{4\eta k}.$$

Moreover, as b_0 is bounded, we deduce that for all $x \in K_T^j$,

$$|b_j^i(x) - b_0^i(x)| \leq \frac{8v}{c_\eta}e^{4\eta k}\left\|b_0^i\right\|_\infty\sqrt{\frac{\log T}{T}} + \frac{8e^{4\eta k}}{c_\eta}\tilde{I}^i[d_T^j](x). \tag{30}$$

We therefore need to evaluate $\tilde{I}^i[d_T^j](x) = \tilde{I}_1^i[d_T^j](x) + \tilde{I}_2^i[d_T^j](x)$ on K_T^j. As

$$\left\|\frac{\partial d_T^j}{\partial x_j}\right\|_\infty \leq \frac{v}{H_j}\sqrt{\frac{\log T}{T}}, \tag{31}$$

it clearly follows that

$$\tilde{I}_1^i[d_T]^j(x) \leq kv\sqrt{\frac{\log T}{T}}\left(\frac{1}{H_1} + \frac{1}{H_2}\right). \tag{32}$$

Regarding $\tilde{I}_2^i[d_T^j](x)$, we can act exactly as in the third point of Proposition 4 of [3]. As $x \in K_T^j$, $x_i \in [x_{j,i} - H_i, x_{j,i} + H_i]$ for $i = 1, 2$. Therefore, using also the definition of d_T^j, the first integral is between $x_{j,i} - H_i$ and x_i. We enlarge the domain of integration to $[x_{j,i} - H_i, x_{j,i} + H_i]$ and then, appealing to (29) and (31) and the fact that the intensity of the jumps is finite, we get

$$
\begin{aligned}
|\tilde{I}_2^i[d_T^j](x)| &\le \int_{x_{j,i}-H_i}^{x_{j,i}+H_i} \int_{\mathbb{R}^2} |d_T^j(\tilde{w}_i) - d_T^j(\tilde{w}_{i-1}) + (\gamma \cdot z)_i \frac{\partial}{\partial x_i} d_T^j(w_i)| F(z) dz dw \\
&\le 2 \, (\int_{\mathbb{R}^2} F(z) dz) \int_{x_{j,i}-H_i}^{x_{j,i}+H_i} \left\| d_T^j \right\|_\infty dw \\
&\quad + \int_{x_{j,i}-H_i}^{x_{j,i}+H_i} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} |(\gamma \cdot z)_i| \left\| \frac{\partial d_T^j}{\partial x_i} \right\|_\infty F(z) dz dw \\
&\le c H_i \sqrt{\frac{\log T}{T}} + \frac{c H_i}{H_i} \sqrt{\frac{\log T}{T}},
\end{aligned}
$$

for some $c > 0$. Using this together with (30) and (32) it follows that, for any $x \in K_T^j$,

$$
\begin{aligned}
|b_j(x) - b_0(x)| &\le c \sqrt{\frac{\log T}{T}} + \frac{8 e^{4\eta k}}{c_\eta} k v \sqrt{\frac{\log T}{T}} \left(\frac{1}{H_1} + \frac{1}{H_2} \right) \\
&\quad + c H_i \sqrt{\frac{\log T}{T}} + c \sqrt{\frac{\log T}{T}} \\
&\le \frac{8 e^{4\eta k}}{c_\eta} k v \sqrt{\frac{\log T}{T}} \left(\frac{1}{H_1} + \frac{1}{H_2} \right),
\end{aligned}
$$

where the last inequality is a consequence of the fact that, $\forall i \in \{1, 2\}$, $H_i \to 0$ as $T \to \infty$ and so, for T large enough, all the terms are negligible when compared to the second one. Hence, the three points listed at the beginning of the proof hold true. We deduce that

$$
\begin{aligned}
&\int_{\mathbb{R}^2} |b_0(x) - b_j(x)|^2 f_0(x) dx \\
&\quad = \int_{K_T^j} |b_0(x) - b_j(x)|^2 f_0(x) dx + \int_{K_T^{j\,c}} |b_0(x) - b_j(x)|^2 f_0(x) dx \\
&\quad \le c \, v^2 \frac{\log T}{T} H_1 H_2 + \frac{64 e^{8\eta k}}{c_\eta^2} k^2 v^2 \frac{\log T}{T} \left(\frac{1}{H_1} + \frac{1}{H_2} \right)^2 |K_T^j|.
\end{aligned}
$$

We recall that $|K_T^j| = H_1 H_2$ and that, as $T \to \infty$, $H_i \to 0$. Thus, the first term is negligible compared to the second one. The desired result follows.

References

[1] Applebaum, D. (2009). Lévy processes and stochastic calculus. Cambridge university press.

[2] Amorino, C., Gloter, A. (2021). Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes. Journal of Statistical Planning and Inference 213, 106-129.

[3] Amorino, C. (2020). Minimax rate of estimation for the stationary distribution of jump-processes over anisotropic Hölder classes. arXiv preprint arxiv.org/abs/2011.11994

[4] Amorino, C., Dion, C., Gloter, A., Lemler, S. (2020). On the nonparametric inference of coefficients of self-exciting jump-diffusion. arXiv preprint arXiv:2011.12387.

[5] Banon, G. (1978). Nonparametric identification for diffusion processes, SIAM J. Control Optim. 16, 380395.

[6] Banon, G., Nguyen, H.T. (1981). Recursive estimation in diffusion model, SIAM J. Control Optim. 19, 676685.

[7] Blanke, D. (1996) Estimation de la densité pour des trajectoires non directement observables, Publ. Inst. Statist. Univ. Paris 40, 2136.

[8] Barndorff-Nielsen, O. E., Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. J. R. Stat. Soc., Ser. B, Stat. Methodol., 63, 167-241.

[9] Bosq, D. (1997). Parametric rates of nonparametric estimators and predictors for continuous time processes, Ann. Statist. 25, 9821000.

[10] Bosq, D. (1998). Nonparametric Statistics for Stochastic Processes. (Second edition), Lecture Notes Statist., 110, New York: Springer-Verlag.

[11] Bosq, D. (1998) Minimax rates of density estimators for continuous time processes, Sankhya. Ser. A 60, 1828.

[12] Bosq, D., Davydov, Yu. (1999). Local time and density estimation in continuous time, Math. Methods Statist. 8, 2245.

[13] Bosq, D., Merlevde, F., Peligrad, M. (1999). Asymptotic normality for density kernel estimators in discrete and continuous time. J. Multivar. Anal. 68, 78-95.

[14] Cheze-Payaud, N. (1994). Nonparametric regression and prediction for continuous-time processes, Publ. Inst. Statist. Univ. Paris 38, 3758.

[15] Castellana, J. V., Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes. Stoch. Process Their Appl. 21(2), 179-193.

[16] Chen, Z. Q., Hu, E., Xie, L., Zhang, X. (2017). Heat kernels for non-symmetric diffusion operators with jumps. J Differ Equ., 263(10), 6576-6634.

[17] Comte, F., Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. In Annales de l'IHP Probabilités et statistiques (Vol. 49, No. 2, pp. 569-609).

[18] Comte, F., Merlevède, F. (2005). Super optimal rates for nonparametric density estimation via projection estimators. Stoch. Process Their Appl., 115, 797-826

[19] Delattre, S., Gloter, A., Yoshida, N. (2020). Rate of Estimation for the Stationary Distribution of Stochastic Damping Hamiltonian Systems with Continuous Observations. arXiv preprint arXiv:2001.10423.

[20] Delecroix, M. (1980). Sur lestimation des densités dun processus stationnaire ẽmps continu. Publications de lISUP, XXV, 1-2, 17-39.

[21] Kutoyants, Y.A. (1997). Some problems of nonparametric estimation by observations of ergodic diffusion process, Statist. and Probab. Lett. 32, 311320.

[22] Kutoyants, Y.A. (1998). Efficient density estimation for ergodic diffusion processes, Stat. Inference Stoch. Process. 1, 131155.

[23] Dexheimer, N., Strauch, C., Trottner, L. (2020). Mixing it up: A general framework for Markovian statistics beyond reversibility and the minimax paradigm. arXiv preprint arXiv:2011.00308.

[24] Dion, C., Lemler, S. (2020). Nonparametric drift estimation for diffusions with jumps driven by a Hawkes process. Statistical Inference for Stochastic Processes 23, 489-515.

[25] Ditlevsen, S., Greenwood, P. (2013). The MorrisLecar neuron model embeds a leaky integrate-and-fire model. Journal of Mathematical Biology 67 239-259.

[26] Doukhan, P. (2012). Mixing: properties and examples (Vol. 85). Springer Science and Business Media.

[27] Funke, B., Schmisser, E. (2018). Adaptive nonparametric drift estimation of an integrated jump diffusion process. ESAIM: Probability and Statistics 22, 236-260.

[28] Hasminskii, R. Z. (1980). Stability of differential equations. Germantown, MD: Sijthoff and Noordhoff.

[29] Kou, S.G. (2002). A Jump-Diffusion Model for Option Pricing. Management Science, 48, 1086-1101.

[30] Lamberton, D., Pages, G., (2002). Recursive computation of the invariant distribution of a diffusion. Bernoulli 8(3), pp.367-405.

[31] Leblanc, F. (1997). Density estimation for a class of continuous time processes, Math. Methods Statist. 6, 171199.

[32] Merton, R.C. (1976). Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3, 125-144.

[33] Nguyen, H. T. (1979). Density estimation in a continuous-time Markov processes. Ann. Statist. 7, 341-348.

[34] Panloup, F. (2008). Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. The Annals of Applied Probability 18(2), 379-426.

[35] Schmisser, E. (2019). Non parametric estimation of the diffusion coefficients of a diffusion with jumps. Stochastic Processes and their Applications 129(12), 5364-5405.

[36] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. The Annals of Statistics 46(6B), 3451-3480.

[37] Tsybakov, A. B. (2008). Introduction to nonparametric estimation. Springer Science and Business Media.

[38] Veretennikov, A. Yu. (1999). On Castellana-Leadbetters condition for diffusion density estimation. Stat. Inference Stoch. Process. 2, 1, 1-9.

[39] Van Zanten, H. (2001). Rates of convergence and asymptotic normality of kernel estimators for ergodic diffusion processes. Nonparametric Statist. 13 (6), 833-850.